

Lawrence Berkeley National Laboratory

Lawrence Berkeley National Laboratory

Title

DOE'S ENERGY DATA BASE (EDB) VERSUS OTHER ENERGY-RELATED
DATA BASES: A COMPARATIVE ANALYSIS

Permalink

<https://escholarship.org/uc/item/54v8s8st>

Author

Robinson, J.

Publication Date

1981-02-01

Peer reviewed

CONF-801195--1



Lawrence Berkeley Laboratory

UNIVERSITY OF CALIFORNIA

Engineering & Technical Services Division

MASTER

Presented at Online '80 Conference, San Francisco, CA,
November 13, 1980

DOE'S ENERGY DATA BASE (EDB) VERSUS OTHER ENERGY-RELATED
DATA BASES: A COMPARATIVE ANALYSIS

Jo Robinson and Margaret Hu

February 1981



DOE's Energy Data Base (EDB) Versus
Other Energy-Related Data Bases:
A Comparative Analysis

Paper Presented at Online-80 in San Francisco, November 13,
1980

by Jo Robinson and Margaret Hu
Lawrence Berkeley Laboratory

1. INTRODUCTION

The release of the DOE Energy Data Base to commercial services in 1980 raised an immediate question in the minds of some searchers: how does this new data base fit into the spectrum of already-available data bases? Because the authors have been closely associated with the Department of Energy's RECON system and its data bases for several years, as trainers and as editor of the DOE/RECON Newsletter, the question was of great interest and we decided to investigate it.

2. BACKGROUND

2.1 The Energy Data Base: History

The Energy Data Base is produced by the DOE Technical Information Center in Oak Ridge. The Technical Information Center was established in 1946 to manage the scientific and technical information generated by the Atomic Energy Commission.

When the TIC became part of the Energy Research and Development Administration, and then of the Department of Energy, the scope of the information covered expanded from nuclear energy information to all scientific and technical information in energy. The EDB now contains over 620,000 citations. Taken together with its predecessor, Nuclear Science Abstracts, it contains over a million citations.

2.2 EDB sources

TIC processes 20,000 DOE reports in a typical 12-month period, in addition to indexing and abstracting information from primary journals, conference proceedings, patents, books, and dissertations.

To supplement its own efforts, TIC contracts with other abstracting and indexing services for coverage of specified material in the data base. Magnetic tape suppliers to the DOE data base are shown in Figure 1, and include (starting from the top) the American Institute of Physics (for in-scope physics information); INIS (for international nuclear information; US nuclear information is supplied to INIS by DOE/TIC); the International Energy Agency (for Coal Abstracts and biomass information); the Federal Republic of Germany; Sweden and other Nordic countries. Further domestic tape suppliers include TULSA, for Petroleum Abstracts; NASA for International Aerospace Abstracts and Scientific and Technical Aerospace Reports, IFI Plenum for patents; the Franklin Institute Research Laboratories for solar information; Government Reports Announcements for NTIS material (DOE/TIC supplies to NTIS abstracts of all DOE sponsored reports); and Engineering Index for coverage of a selected list of journals.

2.3 EDB subject coverage

Broadly stated, the scope of EDB is all current information in DOE's fields of interest. The general subject areas can be grouped by subject category, as: energy sources, including coal and coal products, petroleum, natural gas, oil shales and tar sands, nuclear fuels, fusion fuels, hydrogen, other synthetic and natural fuels, hydro energy, solar energy, geothermal energy, tidal power, wind energy; energy storage, conversion, consumption, conservation, management; nuclear power plants and nuclear reactor technology; environmental and biomedical studies with reference to energy technology and basic sciences as they are directly related to energy research.

2.4 Search Tools for EDB

There are a number of search tools for the EDB. The most important for searching are the DOE Thesaurus and the Energy Categories, listed below. Both are a must for anyone who will need to search energy topics frequently.

The Thesaurus is continually updated, and is reissued annually. The controlled vocabulary it contains is excellent, and together with a powerful hierarchical system provides for comprehensive or specific retrieval at the

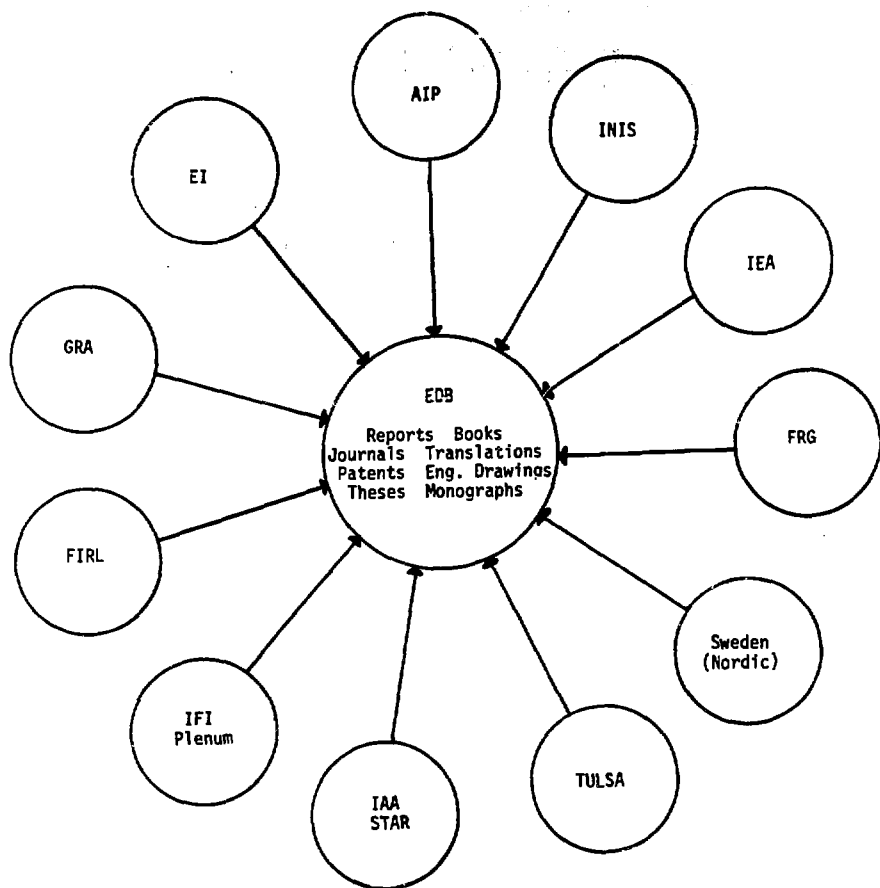


FIGURE 1

searcher's discretion.

The Categories are reissued periodically. They too, by their hierarchical structure, help to make relevant retrieval possible.

3. OBJECTIVES

Knowing that the aim of the Energy Data Base is to be a comprehensive source of energy information, and knowing too that the subject matter covered by the data base has exploded over the past several years, while manpower available to process the information has stayed still, the authors wished to test and compare the EDB with other data bases covering energy information, measuring overlap, coverage of non-US and older material, types of material covered, and currency.

DOE/TIC-7000-R4, Energy Information Data Base: Subject Thesaurus, TIC, September 1977. Available from the National Technical Information Service, Springfield, VA 22161. \$22.50

DOE/TIC-4584-R4, Energy Information Data Base: Energy Categories, TIC, March 1980. Available from the National Technical Information Service, Springfield, VA 22161. \$7.00

4. APPROACH

4.1 Choice of topic

Three search topics were chosen for this investigation. Of course, in a data base as large as EDB and a subject as diverse as energy, no small number of topics can begin to address the coverage in different areas. Nevertheless, because of budget restrictions and because the analysis was done by hand rather than by computer, a search strategy approach was used rather than any other means of checking coverage and overlap.

The topics were chosen from the main areas of synfuels (methane production from agricultural wastes); nuclear energy and the environment (risks of transporting radioactive wastes) and oil and gas statistical information (reserve estimates of petroleum and natural gas in the United States).

4.2 Preparation of search strategies

The search strategies used were aimed at relevance rather than recall, but were treated broadly; relevance judgements were not quantified, but in general the strategies worked rather well. Problems will be discussed with each search.

4.3 Choice of data base

For each search topic, as many data bases as seemed likely candidates were used. Had these searches been done for a real patron, it is unlikely that so many data bases would have been used. Nevertheless, the results will be useful in making future choices of data bases to search.

It was decided to limit the scope of the searches by excluding the research-in-progress and newspaper data bases, though both might be expected to have information on the search topic.

4.4 Analysis

After executing the searches the results were analyzed manually: citations were labeled, copied, and sorted. Duplicates, publication dates, document types and languages were tallied; results are given in Tables 1-6.

For the duplication-checking, in order to avoid discrepancies due to variations in currency of input, a 5-

year span, 1975-1979, was used. Duplicates were found as the various searches were merged by author, title and year. Only exact duplicates were counted; those in different forms such as preprint vs. paper in conference proceedings were not counted.

5. THE SEARCHES

5.1 Methane (or Biogas) Production from Agricultural Wastes

This topic cuts across several disciplines: as an alternative energy source, or synfuel it belongs to energy; as a use of agricultural wastes to agriculture; as a bioconversion process, to biology and chemistry.

Data bases searched included EDB, Chem Abstracts, Agricola, NTIS (because of its two-way communication with EDB), Energyline, CAB, BIOSIS, Scisearch, Enviroline, Conference Papers Index, GPO and Dissertation Abstracts.

5.1.1 Search Strategy

EDB. This search topic is one that had been previously used by one of the authors for a comparison; at that time it had become clear that the EDB strategy should not be too restrictive, or a reasonable comparison would not result. For this reason the search strategy used was the two-facet strategy shown in Figure 2.

Note that the search could be approached very simply, because of the strong, hierarchical vocabulary available on EDB. The terms METHANE and BIOGAS PROCESS, from the EDB Thesaurus, were OR'd with the title-word BIOGAS (on RECON, a prefixed field, TL=BIOGAS), and with the term ANAEROBIC DIGESTION, which is the bioconversion process by which methane is produced.

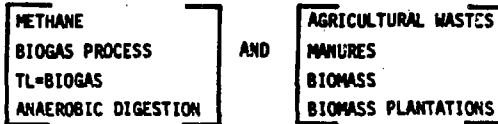
The results were ANDed with the terms AGRICULTURAL WASTES, BIOMASS and BIOMASS PLANTATIONS. AGRICULTURAL WASTES is a hierarchy containing narrower terms MANURES and BAGASSE (MANURES is shown in the strategy only because it was carried over to the strategy for the other data bases). Both narrower terms were retrieved when the broader term was entered.

The simple strategy shown compared favorably with that used on the other data bases; because EDB is well indexed it was not necessary to add the long list of animals and crops

METHANE
FROM
AGRICULTURAL WASTES

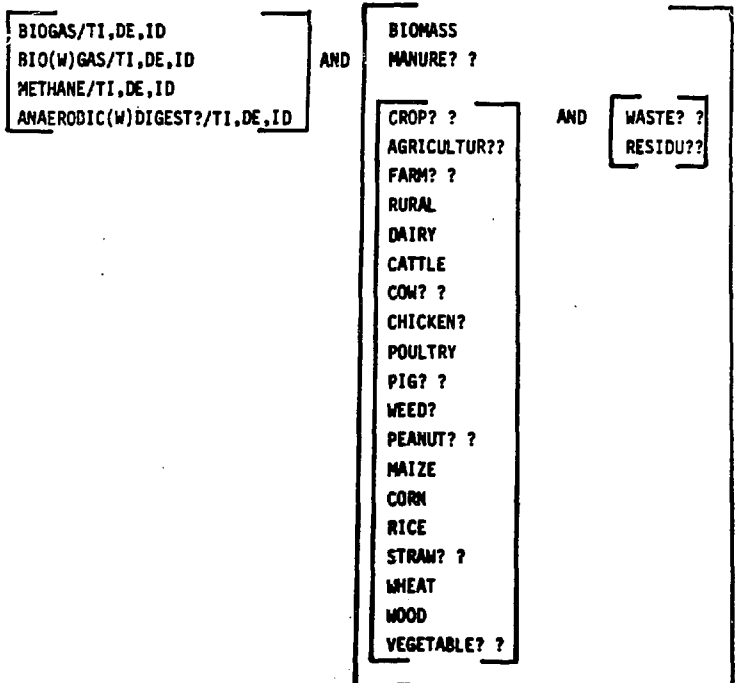
DOE EDB Search Strategy

FIGURE 2



Search Strategy for Other Data Bases

FIGURE 3



whose residues could be used for methane production, as was done for the free-text version of the comparison search.

The EDB strategy retrieved 887 citations, of which 535 were in the sample range, 1975-1979.

Other data bases. Compare the EDB strategy with that for the other data bases. (Figure 3) Note that the main concept, methane or biogas, was qualified so as to retrieve only from the title and index terms (controlled or free). This had the result of focussing this facet of the search so that it had some importance in every citation retrieved. The particular kind of waste or residue, however, was not qualified because retrieval from the abstract would be acceptable for this facet.

5.1.2 Problems

Size. The biggest problem in this search was its size. The authors were most grateful for the assistance of colleagues from the Information Research Group in labeling and tallying citations. 1507 citations were checked!

Dissimilar strategies. It was expected that variations in search strategy would result in the missing of some duplicates, especially since an indexing term approach, used on EDB, was contrasted with a partly free-term approach on the other data bases.

Control search. For this reason, since EDB had in the meantime become available in a free-text environment on a commercial search service, a control search using the looser strategy employed for the other data bases was run. "Unique" citations and non-EDB duplicates were rechecked against this search. It retrieved 951 citations of which 696 were checked.

5.1.3 Results

Table 1 shows the results of the citation-by-citation duplication check. The data bases are listed from the top, by number of citations retrieved. Figures shown for EDB reflect the control search done in free-text mode on a commercial search service.

Reading across the top row, it may be seen that out of 951 EDB citations, 696 (1975-1979) were checked; that 264, or 38% of those checked, were duplicates with one or more other data bases, and that 432, or 62% of those checked, were unique to EDB.

Following across, the table shows that EDB had 6

METHANE FROM TOTAL
AGRICULTURAL CITS/
TOTAL TOTAL
WASTES CHECKED % %

DUPLICATE CITATIONS WITH OTHER DATA BASES

WASTES	TOTAL CHECKED	X	X	EDB	ChemAbs	Agric	NTIS	Energy1	CAB	Biosis	SciSearch	Envir	Conf P	GPO	Diss															
EDB	951	696	264	381	432	6	11	63	91	26	41	133	191	61	91	6	11	8	11	6	11	4	11	1	01	11	21	2	-	
Chem Abs	222	170	99	581	71	421	63	371		27	161	1	91	18	111	9	51	11	61	12	41	5	31	2	11	2	11	2	11	
Agricola	213	148	60	401	601	26	111	27	111	13	71	1	11	12	21	8	51	11	71	12	41	1	11	2	11	11			11	
NTIS	187	175	143	821	111	133	761	15	91	1	11			2	11			1	11								9	51		
Energyline	128	111	71	641	40	561	61	551	18	161	12	111	2	11		3	31	1	11	4	41	1	11							
CAB	88	73	18	241	55	741	6	41	9	121	8	141		3	41	2	31	2	31	4	51									
Biosis	66	37	18	491	19	511	8	221	11	301	11	301	1	31	31	2	51			6	161			1		31				
SciSearch	44	35	24	691	11	311	6	171	5	401	12	341		4	111	4	111	6	171			1	31	2	61					
Enviroline	42	21	9	421	12	561	4	191	2	241	1	51		1	51				171	1	51									
Conf Papers	32	23	4	171	19	831	1	41	2	91	2	91						1	41	2	61									
GPO	15	15	13	871	2	131	71	731	131				9	601												1		71		
Biosis Abs	4	3	2	671	1	331	2	671	2	671														1	251					
1967-1980																														
TOTAL	1993	1507	725	782																										

TABLE 1

duplicates with itself, 63 with Chem Abstracts, 26 with Agricola, and so on. Percentages on this line refer to the total EDB citations checked; thus, 9% of the EDB citations were also found on Chem Abstracts. Note that duplicates between pairs of data bases (figures and percentages across each row, from column 4) may add up to more than the total in column 2; this is because of multiple duplication.

Moving to the next line, 170 out of 222 Chem Abstracts fell within the 1975-1979 period and were checked; 99 duplicates were identified, leaving 71 unique citations. Percentages again refer to the total of this data base checked, so moving again to the 63 EDB-Chem Abstracts duplicates, 37% refers to the percentage of Chem Abstracts citations duplicated on EDB.

With the postulate that for an energy search, EDB should be the "first stop", let us turn our attention to the vertical column labeled EDB. Here may be found the clearest indication of the relative usefulness of each data base, for this topic, in terms of duplication with EDB. A large number of unique citations (column 3) coupled with a low percentage of overlap (column 4) indicates a good source.

The Chem Abstracts 37% duplication with EDB means that the remaining citations (63%) were not found on EDB by the strategy used. 148 out of 213 Agricola citations were checked, and only 18% were duplicates with EDB. Thus Agricola as well as Chem Abstracts can be seen to be useful sources for the topic, Methane Production from Agricultural Wastes.

NTIS on the other hand produced many duplicates with EDB (as expected), with 133 or 76% duplicates. Normally one would not need to search NTIS for energy topics, since not only does DOE/TIC send EDB-announced reports to NTIS, but TIC also extracts other information from NTIS that falls within the subject scope of EDB.

Energyline had a substantial amount of duplication with EDB, with a smaller number of citations. Some of the unique items found were from the popular literature, or from smaller conferences.

CAB showed only a small amount of overlap with EDB, 6 citations, or 8% of the retrieval. CAB, with its significant international coverage, is an excellent source for this agricultural-energy topic of third-world importance.

Biosis contributed 19 unique citations to this search; 22% of the items overlapped with EDB, while 30% each overlapped with Chem Abstracts and Agricola. Although not shown in the search strategy diagram a cross code was used in the Biosis strategy.

Scisearch contributed 11 unique citations and 24 that were duplicates, using the same search strategy as on the other commercial data bases. Normally one might expect to make a title-only search strategy a little simpler than one intended to retrieve from indexing and abstracts; however, Scisearch performed well with the one shown.

Of the other data bases, the Conference Papers Index should be mentioned, since it retrieved a modest, but ostensibly "unique" number of retrieved citations. In fact, it is difficult to compare conference citations; a preprint is a different piece of information from a paper in a published proceedings, and these were treated as non-duplicates. Since the Conference Papers citations were usually in the form of preprints, the data base had a high "unique" rating; it was clear that many of the seemingly unique citations were in fact preprint versions of articles also found in a published version.

Turning to Table 2, note that the "total citations" column shows 887 for EDB, with a total of 1929 citations on all data bases. This table was not recalculated on the basis of the control search and thus does not show the same totals as Table 1. The publication dates shown in Table 2 illustrate the fact that the Energy Data Base contains subject areas in which the coverage goes far back -- in this case to 1939, but in other cases to the early 1900's. This was because of a retrospective indexing effort begun in 1974 with the creation of the Energy Research and Development Administration, predecessor of the Department of Energy.

As to currency, EDB had the largest number of citations with publication date in 1980; but Scisearch, as always, won the prize for currency with 20% of the retrieved items from 1980.

Foreign coverage in EDB appeared to be good, both foreign-language coverage, and English but non-US coverage. EDB, Chem Abstracts, CAB and Biosis retrieved about one quarter foreign-language articles, while Chem Abstracts, Agricola, CAB, Biosis, Scisearch and Enviroline all retrieved over 20% English but non-US citations, on this search.

By type of document the data bases did not show much difference, except for obvious differences such as NTIS primarily covering reports, and Scisearch journal articles. EDB showed good coverage of conferences, along with Energyline and Conference Papers Index.

METHODS FROM AGRICULTURAL MATERIALS	TOTAL CITIS	Publication Dates					Language		Document Types										
		Earliest Year of Public- ation	Pre-1965-1980				English	Foreign	Report	Journal	Conf	Patent	Book	Other					
			Pre-1965	1965-1974	1975-1979	1980													
74-80	987	1939	250	281	85	535	17	21	530	135	222	221	369	193	20	21	89	12	11
Chem Abs	227	1964	1	40	170	11	51	118	49	151	55	20	140	37	20	3	101	2	2
Agri Coll	213	1951	10	50	148	5	24	130	59	223	251	53	91	631	171	51	18	11	11
NTIS	167	1963	2	10	175	708	21	163	611	228	171	100	250	491	7	111	111	7	38
Energyline	128							105	23		31		22	988	7	48			
CAN	89		2	11	73	3	31	23	41	181		16	171	581	2	441		1	11
Sci Search	66		25	37	4	31	61	28	21	461	241		16	62	201	51	9	1	101
Enviroline	44	1976		381	561	9	61	411	321	261			46	781	10	8			
Conf Papers	42	1970	20	21	1	201	21	32	10	251	41		1	35	811	21			
GPO	15		2	23	7	221	21	32	10	241			6	28	611	6	1	21	21
Diss Abs	4							29	3	161			141	611	32	141			
1981-1980	4							15		161			15						
TOTAL 1989			265	141	131	1346	60	1229	351	349	111	553	357	371	160	134	35	35	21

1634-00018

TABLE 2

5.2 Risk of Transporting Radioactive Wastes

Nuclear energy has long been a controversial subject, one aspect of which has been the transportation of radioactive materials. This search topic was restricted to the transportation of radioactive wastes. This included high or low-level radioactive wastes, and spent fuels in various forms. Transportation could be by rail, by truck, or by sea. We were interested in risk assessment, accident or safety analysis, and hazards in the transportation process. Other safety aspects such as sabotage or theft were not included.

Eighteen data bases were used. Listed in order of retrieval, they were: EDB, NSA, NTIS, TRIS, Bioline, INSPEC, NSIC, Energyline, Chem Abstracts, Compendex, CIS, ASI, Federal Index, Scisearch, Magazine Index, ISMFC, PAIS, Social Scisearch. Of these, NSA and NSIC were available only on DOE/RECON. Others were or were becoming available on commercial online systems such as DIALOG, ORBIT or BRS.

5.2.1 Search Strategy

EDB. In the search strategy for EDB, only thesaurus terms were used. The index terms are grouped by concept in Figure 4. EDB has very specific terms on nuclear energy. Besides RADIOACTIVE WASTES, there are terms on RADIOACTIVE WASTE DISPOSAL, RADIOACTIVE WASTE MANAGEMENT, and many terms on SPENT FUELS. There is also a subject category NC=052002 on transportation & handling of radioactive wastes.

As mentioned above, there is a hierarchical structure in the thesaurus terms. Looking at the second concept, the term TRANSPORT includes narrower terms (NTs) such as LAND TRANSPORT, RAIL TRANSPORT, and MARINE TRANSPORT.

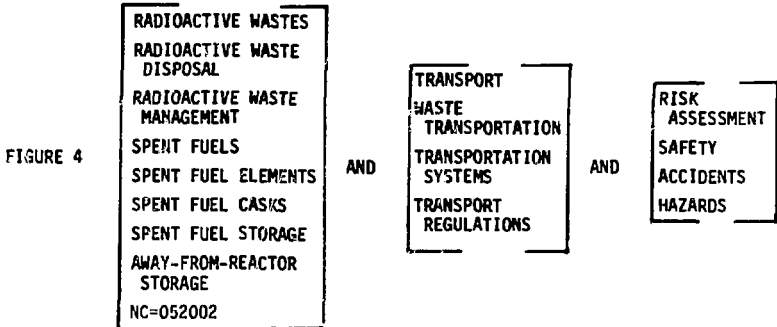
In the third concept, risk, the term HAZARDS includes the NTs HEALTH HAZARDS, RADIATION HAZARDS and FIRE HAZARDS.

NSA. Nuclear Science Abstracts was the forerunner of EDB, covering nuclear information. It was published by the Technical Information Center, then under the USAEC, from 1948-76, but only the last ten years have been put online, i.e., vol. 21-33, from 1967-June 76. Although NSA has been discontinued, the information is still significant and useful for nuclear reactor technology and related areas. In fact, NSA is the second most frequently used data base on the DOE/RECON system after EDB.

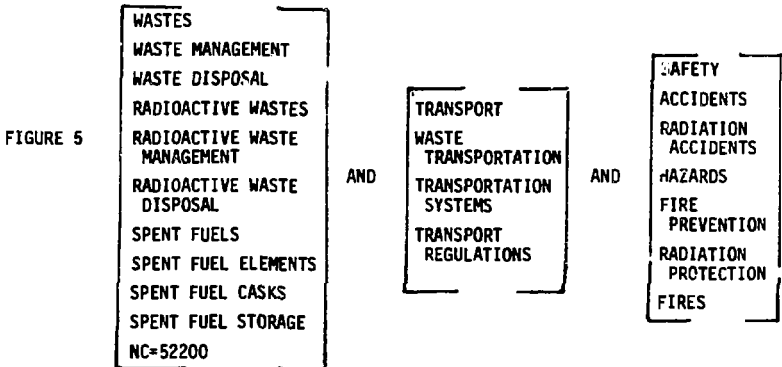
The search strategy employed on NSA was similar to that for EDB (Figure 5). However, more terms were used because NSA is older and has gone through several changes in its

RISKS OF TRANSPORTING RADIOACTIVE WASTES

Search Strategy for DOE Energy Data Base



Search Strategy for Nuclear Science Abstracts



thesaurus development. Also, there is no hierarchy in the thesaurus terms in NSA as loaded on RECON. Therefore, terms like FIRES, and RADIATION PROTECTION needed to be used with the term HAZARDS.

NSIC. The NSIC data base was produced by the Nuclear Safety Information Center at the Oak Ridge National Laboratory. The data base contains very specialized information on nuclear reactor safety. The Center is supported about 85% by the Nuclear Regulatory Commission (NRC) and 15% by DOE. This is reflected in its content and coverage. About two-thirds of the items in NSIC are documents from NRC, which includes many unusual items not found on other data bases, e.g., licensee event reports, safety analysis reports, correspondence, or questions and answers between NRC and the nuclear facilities.

The NSIC data base goes back to 1963, and has its own vocabulary. The search strategy used was very similar to that for EDB (Figure 6). The keywords in the third concept show many specific terms on safety and accident analysis since this is the primary concern of the data base.

Other data bases. Figure 7 shows the search strategy used on the other data bases which was somewhat broader than those on EDB, NSA and NSIC. The same strategy was used across all the other data bases rather than using the specific descriptors pertinent to each one.

The basic index including title, indexing, and abstract words was used for this search rather than selecting from just title words or descriptors. This approach was used because waste transport is often discussed in conjunction with waste management, waste disposal or radioactive materials management, and might not be the primary emphasis of the paper.

For the first concept, RADIOACTIVE was selected in the same field as WASTE rather than RADIOACTIVE adjacent to WASTE. This was because of the various forms in which the concept could appear: RADIOACTIVE MATERIALS AND WASTES, NUCLEAR MATERIALS, NUCLEAR WASTES, and INDUSTRIAL WASTE.

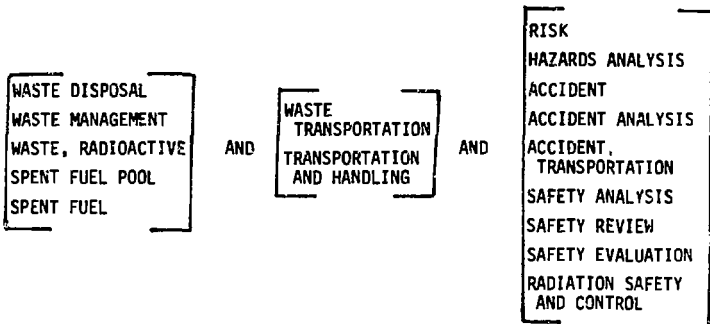
5.2.2 Problems

False drops. Some irrelevant citations were retrieved because of different meanings of the word TRANSPORT, e.g., ENVIRONMENTAL TRANSPORT in the air or water, RADIONUCLIDE TRANSPORT. The term TRANSPORT was legitimate and had to be used.

RISKS OF TRANSPORTING RADIOACTIVE WASTES

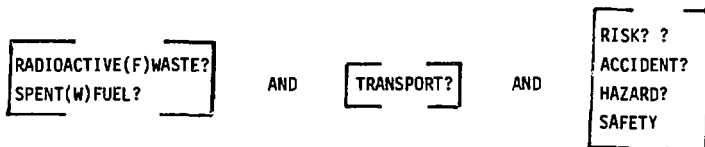
Search Strategy for Nuclear Safety Information Center Data Base

FIGURE 6



Search Strategy for Other Data Bases

FIGURE 7



Another group of less specific items retrieved came from the truncated word: HAZARD? ?. In some titles and in a few data bases, the term HAZARDOUS MATERIALS is synonymous with RADIOACTIVE MATERIALS AND WASTES. In such cases, the retrieval was effectively made on two concepts rather than on three, i.e., TRANSPORT and HAZARDOUS WASTES. The third concept RISK, was therefore left out. On the other hand, the term "HAZARD" could not have been omitted from the search because some references were on health hazards, radioactive hazards or fire hazards in waste transportation.

Dissimilar Strategies. The EDB search on RECON resulted in a retrieval of 255 citations of which 223 fell into the 1975-1979 period. This search was restrictive because the "risk" concept was expressed in index terms on EDB, but by free terms on other data bases.

Control search. Following the same approach as for the methane search, we also ran the "other data base" strategy shown in Figure 7. With this strategy 277 citations were found in the 1975-1979 period. The overlap figures in Table 3 reflect this broad search.

5.2.3 Results

In Table 3, EDB & NSA are listed at the top, followed by the other data bases listed in order of retrieval. In EDB, of 277 citations, 110 were duplicates and 167 items were unique. In NSA, 230 citations were retrieved. Of these, only 30 fell into the period checked. Twenty of these were duplicates and ten items were unique. EDB and NSA taken together had 548 citations, with 307 in the period checked.

Among the other data bases, NTIS retrieved a large number of items, 268 over a 17 year period. TRIS was next, with 127 items. The third column in Table 3, TOTAL DUPS, shows that both NTIS and TRIS have many duplicates with other data bases including each other.

Looking at the fourth column, TOTAL UNIQUE, EDB had 167 unique items, far more than any other data base. The fifth column of Table 3 under EDB shows the duplicates between EDB and other data bases. A few data bases did not have overlaps with EDB on this topic because their contents were more specialized, e.g., items from the Federal Index, or Congressional Information Service. Others such as ENERGYLINE, ENVIRONMENTAL INDEX, or MAGAZINE INDEX contain more popular literature which does not fall into the technical scope of EDB.

In Table 4 figures for EDB reflect the narrower DOE/RECON search for a total of 255 citations. In the first section on years of coverage, EDB did very well on currency:

Risks of Transporting Radioactive Waste	Earliest Year of Public- ation	Pre- 1965 & n.d.	1965- 1974	1975- 1979	1980	English		Foreign		Report		Patent	Book	Other
						US	non-US	US	non-US	Journal	Conf			
EDR 74-57	255		20 44	224 848	11 41	166 654	43 174	46 181		112 448	45 174	76 388	10 48	12 54
NSA 67-76	230	9 41	191	30		174 43	43 101	13 41		147 641	52 23 2	39 175	6 32	4 8 78
NTIS 64-80	268	2 17	53 221	197 73	11 41	239 99	15 64	14 54		259 978	1 31 181	47 181		4 18
TRIS 70-80	127	3 21	10 41	114		113 89	13 1	1 41		119 948	3 4 31	4 31		2 1 48
Enviroline 71-80	90	3 51	22 241	64 1	11	80 49	9 101	1 11		38 421	27 301	22 44	1 11	3 2 21
INSPEC 69-80	73		10 741	62 1	11	42 58	22 301	9 121		5 60 321	9 121			1 2 48
NSIC 63-80	72	8 111	38 531	26 751	1	53 17	241 38	2 38		47 651	7 9 71	6 41	6 9	9 12 58
Energyline 71-80	65	1 21	15 231	49		44 64	15 235	6 91		19 29	16 44			2 38
Chem AB 67-80	63		21 331	41 1	21	34 54	9 141	20 381		23 33 511	19 241	3 4 41	3 4 41	5 9 48
Compendex 70-80	52		22 421	30 511		41 741	9 171	2 41		2 361	23 441	21 441	2 3 41	39 97 58
CIS 70-80	40		11 281	27 681	2 51	40 1001				15 941				1 41
AST 73-80	16		4 251	11 621	1 61	16 1001								10 1001
Fed Index 78-80	10		251 8	621 2	201	10 1001								
SciSearch 74-80	5			5 1001		5 1001				5 1001				
Magazine In 77-80	2			1 501	501	2 1001				2 1001				
ISREC 73-80	1			1 1001	1	1 1001				1 1001				
PAIS 76-80	1			1 1001	1	1 1001				1 1001				
Soc SciSearch 72-80	1			1 1001	1	1 1001				1 1001				
TOTAL	1371	25 41	432 311	891 651	23 21	1062 781	195 141	114 41		787 571	289 211	259 191	20 18	34 21 41

1980-199

TABLE 4

eleven items, or 4% of the EDB retrieval were published in 1980. NTIS also did well on currency: 11 1980 items, or 4% of the NTIS retrieval. These searches were done in September-October, 1980.

In the second section of Table 4, Chem Abstracts covered the highest percent of foreign materials - 32%, but EDB had the highest number, 46. Likewise, EDB showed very good non-US English coverage: 43 items or 17%.

In the third section, EDB's coverage of document types was very close to the total average. EDB also had the highest retrieval of patents.

5.3 Reserve Estimates

The third topic for this study was estimates of US petroleum and natural gas reserves published in 1979 and 1980.

Although petroleum and natural gas are not areas in which EDB tries to be comprehensive (since other comprehensive sources, TULSA and API, are available), a topic on oil and gas was chosen to see how EDB fared in comparison with other data bases in this area.

Fourteen data bases available on DIALOG, ORBIT, or BRS were used in this search. They were: EDB, PROMT, TULSA, ASI, PAIS, NTIS, ABI/INFORM, COMPENDEX, FEDEX, GEOARCHIVE, GEOREF, ENERGYLINE, MANAGEMENT CONTENTS, and ECONOMIC ABSTRACTS. We did not search in any non-bibliographic or numeric data bases that may be available on other systems.

5.3.1 Search Strategy

EDB. In this search, a special feature of EDB, not found on most other data bases was used. It is called major descriptor pairs (Figure 8).

While indexing an article, the indexers at DOE/TIC not only assign descriptors for the major aspects of the article, but also link these with other descriptors to show a definite relationship between them. For example, on DOE/RECON if one selects MD=PETROLEUM/RESERVES, the citations retrieved will have as the major aspect of the paper, the RESERVES of PETROLEUM, not any other energy resources that may be mentioned in the same article. This pairing or linking eliminates any false drops.

Citations retrieved from the major descriptor pairs were ANDed with USA and the years 1979 and 80. The descriptor USA included all states as its narrower terms. With this strategy, 38 very relevant citations were retrieved, but of course this search was by no means comprehensive.

Therefore a less specific, but good version of the previous strategy was used, which is shown in Figure 9. One characteristic of the major descriptors is that they can be selected in pairs as well as separately. In this strategy, major descriptors were selected for the terms for oil and gas. These were ANDed with the regular descriptors on RESERVES, RESOURCE ASSESSMENT, etc. With this strategy, 167 citations were retrieved.

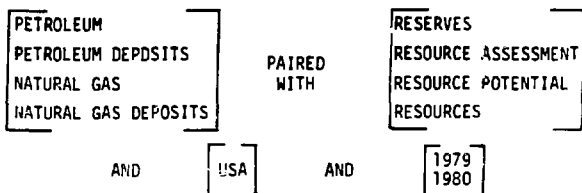
FEDEX. The Federal Energy Data Index, available on DOE/RECON and BRS, is produced by DOE's Energy Information Administration to cover statistical information from all EIA

PETROLEUM & NATURAL GAS
RESERVE ESTIMATES
FOR THE US

Search Strategy for DOE Energy Data Base

Most Specific Version: Major Descriptor Pairs

FIGURE 8



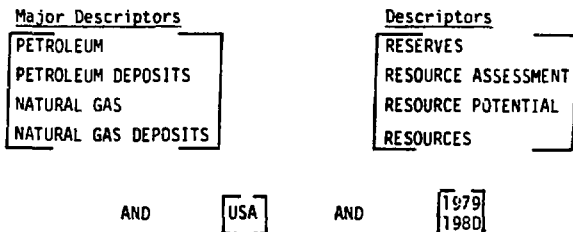
Retrieves 38 Citations

Search Strategy for DOE Energy Data Base

Less Specific, but Good Version:

Major Descriptors AND Regular Descriptors

FIGURE 9



Retrieves 167 Citations

publications. The structure of the data base and the indexing and abstracting are similar to EDB. Therefore the search strategy for FEDEX was essentially the same.

The unique feature of FEDEX is that it includes citations on both the document level and on the table-and-graph level. The citation in Figure 10 is indexed to the table level. The data locator field tells from which report number the citation comes, with the paging and the table number. If one traces by the report number, DOE/EIA-0196, one will find the actual report.

Figure 11 shows the actual items which were indexed and abstracted: Figure 27 and Table 27 on page 58-59. However, it must be emphasized that FEDEX does not give the actual tables and graphs, but only references to them.

Other data bases. One search strategy (Figure 12) was used for the other ten data bases except for TULSA and PROMT. The first concept on oil and gas was limited to title words, descriptors, identifiers or auxiliary terms (the suffix GS is only applicable to GEOARCHIVE). The other concepts were selected from the basic index. Since TULSA is very well indexed, its thesaurus terms were used in that strategy (not shown). As for PROMT, specific codes pertinent to that data base were used.

5.3.2 Problems

Year of publication. It was not possible to retrieve by year of publication on some data bases. When necessary, we selected the year of entry in the data base and then sorted out those published in 1979 and 80 manually.

False drops. There were some false drops from the word RESOURCES when it was imbedded in another phrase, such as WATER RESOURCES affected by oil drilling. Another kind of false drops came from articles on production from natural GAS RESOURCES, rather than RESERVE estimates of natural GAS resources.

Incomplete citations. PROMT presented some problems because it did not give titles or authors for journal sources. This made it difficult to compare with other data bases.

Dissimilar Strategies. On this search we again found ourselves comparing unequal strategies. With the EDB version we used features that help to produce very relevant citations and avoid false drops.

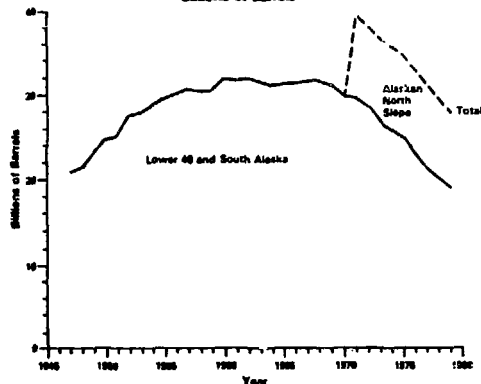
Control search. For better results on duplicate checking, we reran the "other data bases" search for this topic

PAGE 1

<ACCESSION NO.> B0*0002759
<DATA LOCATOR> DOE/EIA--0196 Table 27, figure 27, pages 58-59
<TITLE> Proved crude oil reserves
<DATA DATE> 1947-1979
<DATA AGGREGATE> year;US
<ABSTRACT> Proved crude oil reserves in the Lower-48 States and South Alaska are shown (in billion barrels) for each year from 1947 through 1979. Data are also given for the Lower-48 States and South Alaska plus Alaskan North Slope beginning with the year 1970. An accompanying cumulative line graph illustrates data and includes totals.
<DESCRIPTORS> ALASKA;ALASKAN NORTH SLOPE; T;PETROLEUM; T;PETROLEUM DEPOSITS; T2;RESERVES; Q1;Q2
<DATA SOURCE> ''Reserves of Crude Oil, Natural Gas Liquids and Natural Gas in the United States and Canada as of December 31, 1978''
<DSN> B0;002759
<UPSTED DESC> ALASKA;ENERGY SOURCES;FOSSIL FUELS;FUELS;GEOLOGIC-DEPOSITS; NORTH AMERICA;PACIFIC NORTHWEST REGION;RESOURCES;USA
<LITERATURE TYPE> B

FIGURE 10

Figure 27
Proved Crude Oil Reserves
billions of barrels



Sources: *Reserves of Crude Oil, Natural Gas Liquids and Natural Gas in the United States and Canada as of December 31, 1978*, pgs. 24, 27.

These figures were derived from Supplementary Table E in the back of this publication.

Proved crude oil reserves as of the beginning of the year were used. Alaskan North Slope (Prudhoe Bay) can be derived by subtracting Lower 48 and South Alaska from Total (Lower 48 and South Alaska plus Alaskan North Slope).

Table 27 Proved Crude Oil Reserves
(billions of barrels)

Year	Lower-48 and South Alaska	Lower-48 and South Alaska Plus Alaskan North Slope
1947	20.874	--
1948	21.458	--
1949	23.220	--
1950	24.649	--
1951	25.268	--
1952	27.468	--
1953	27.961	--
1954	28.594	--
1955	29.561	--
1956	29.012	--
1957	30.435	--
1958	30.390	--
1959	30.536	--
1960	31.719	--
1961	31.613	--
1962	31.759	--
1963	31.389	--
1964	30.670	--
1965	30.991	--
1966	31.252	--
1967	31.452	--
1968	31.577	--
1969	30.707	--
1970	29.432	29.632
1971	29.641	39.001
1972	28.463	38.063
1973	26.739	36.330
1974	25.700	35.300
1975	24.650	34.250
1976	23.082	32.682
1977	21.542	30.942
1978	20.186	29.486
1979	18.904	27.404

Search Strategy for Other Data Bases

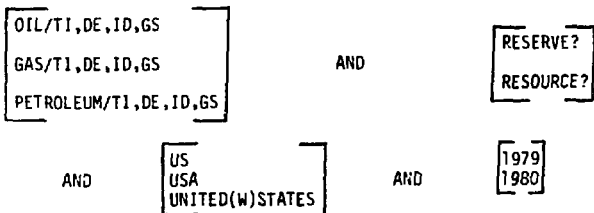


FIGURE 12

also.

5.3.3 Results

With the broader strategy, EDB had by far the largest retrieval, up from the former 167 citations to 647! PROMT had the second, and TULSA the third largest retrieval. In PROMT, over 1/3 of the citations retrieved came from corporate annual reports.

EDB's coverage in this area is substantial. We were surprised that the much larger control search did not reveal much additional duplication with EDB beyond that found initially. In column 3 of Table 5, several data bases show large percentages (though smaller numbers) of unique citations on this topic, indicating that they are good sources covering different kinds of material.

Table 6 shows tabulations on language and document types. In the first section, it is not surprising that most items, 93% of the total, came from the US on a topic of US oil and gas reserves. In the second section, EDB contains a higher percentage of reports, 42%, compared to a higher percentage of journal sources in both PROMT and TULSA, 60% and 64% respectively.

DUPLICATE CITATIONS WITH OTHER DATA BASES

RESERVE ESTIMATES	79-80	Total	Total	Total	EOB	PROCT	TULSA	ASI	PAIS	MTIS	AOI/INF	COMEN	FEDX	GEORG	GEORGE/EMERIT	MOET	ECON								
		Check	Doc	Unlinc													AMS								
EOB	74-81	647	56	592	911	1	18	4	19	8	16	3	2	20	39	2	17	31	3	26	4	19	1	18	
PROCT	72-80	158	10	168		4	31	1	16	4	38							1	18	4	38				
TULSA	65-80	98	15	113		8	21	4	48											4	48	2	1	18	2
ASI	71-80	44	5	49		3	71						1	19					14	5					
PAIS	74-80	46	3	49		2	51		1	18															
MTIS	64-80	32	20	52		20	43																		
AOI/INF	71-80	31	6	37		2	41	1	36	1	36		1	38						2	41		1	36	
COMEN	79-80	27	16	43		17	43	4	158	3	118														
FEDX	77-80	270	18	288		3	118																		
GEORG	71-80	20	8	28		4	201	4	209	4	209														
GEORGE/EMERIT	61-80	17	2	19		1	41		2	121															
MOET	71-80	36	7	43		8	501		1	41															
ECON	74-80	16	2	18		1	71		1	71															
PROCT	74-80	9	3	12		1	118		2	218															
	79-80	1101	147	1248																					
*Citations both to tables within documents and to unique documents.																									
*Based 16 citations are to tables within four different documents.																									
*These 14 table level citations from FEDX occur within three documents cited by ASI.																									
*MOET and GEORGE/EMERIT are digitized by DAME citations in ASI.																									

*Citations made to tables within documents and to unique documents.

These 16 citations are to tables within four different documents.

These 16 table level citations from FEDX occur within three documents cited by ASI.

These 16 citations in FEDX are duplicated by three citations in ASI.

TABLE 5

RESERVE ESTIMATES	TOTAL CITS	Document Types											
		English		Foreign		Report		Journal		Conf	Patent	Book	Other
		US	Non-US										
EDB	167	161	4	2	1	70	42%	49	29%	25	15%	23	14%
74-80		97%	2%										
PROMPT	158	150	5	3		60	38%	95	60%				3
72-80		95%	3%										2%
TULSA	98	84	13	1		7	7%	63	64%	24	25%	2	2
65-80		86%	13%									2%	2%
ASI	44	44				40	91%			1	2%	1	2
73-80		100%										2%	5%
PAIS	40	40				6	15%	28	70%			1	5
76-80		100%										2%	13%
NTIS	32	32				32	100%						
64-80		100%											
ABI/INFORM	31	27	4					31	100%				
71-80		87%	13%										
COMPENDEX	27	20	2	5				20	74%	7	26%	1	
70-80		74%	7%		19%							4%	
FEDEX	27	27				27	100%						
77-80		100%											
GEOARCHIVE	20	17	2	1		1	5%	17	85%	2	10%		
74-80		85%	10%		5%								
GEOREF	17	17						16	94%	8	47%	1	
61-80		100%										6%	
ENERGYLINE	16	16				9	56%	5	31%	3	19%		
71-80		100%											
MGMT CONTENTS	15	15						15	100%				
74-80		100%											
ECON ABS	9	5	3	1				8	89%			1	
74-80		56%	33%		11%							11%	
										</			

TABLE 6

CONCLUSION

This study did not attempt to examine the entire coverage of the Energy Data Base in comparison with other data bases. It did show that EDB has very good, but not complete coverage of energy information in the areas selected. EDB scored high in numbers of citations retrieved, and contained a high number of unique citations. Compared to other data bases, it also did well in currency and foreign coverage. The document types included a good proportion of various types of materials.

To sum up, EDB is a very large, multi-disciplinary data base on energy. It is well indexed and well abstracted. It allows for specificity of searching, e.g. through major descriptor pairs. It has good foreign coverage because of foreign exchange agreements.

Since 1974, when the scope of its parent agency was expanded from predominantly nuclear energy to include many other forms of energy, the DOE Technical Information Center in Oak Ridge has steadily broadened the coverage of the Energy Data Base. Steps are being taken to achieve even greater comprehensiveness of energy - related topics in the future.

It is recommended that the DOE Energy Data Base be searched first on any energy or energy-related topic. Depending on the information needed, other data bases may be searched in addition, to increase comprehensiveness.